

# An analysis of the gender wage gap in Peru in the formal and informal sectors

MSc in Economics, Universitat de Barcelona

## Author

Richar Quispe Cuba

## Advisor

Raul Ramos Lobo

June 18, 2020

## Abstract

Unlike previous studies on gender gaps, this research analyzes wage gaps within the Peruvian labor market considering its segmentation between informal and formal and a sub-segmentation between public and private, generating 4 segments in the labor market. For this purpose, we apply decomposition methods for mean wage and along the wage distribution. The main findings are the following: first, in the informal segment there are wide wage gaps in the upper and lower ends of the distribution against women with a greater emphasis on the informal public segment, and in the informal private segment we found a negative wage gap against women but constant along the distribution. Then, in the formal segment (public and private) the largest gaps were found at the upper end of the distribution, a result that is in line with the "glass ceiling" effect found by the international literature.

**Keywords:** Gender, wage gap, formal, informal, public, private

**JEL classification:** J16, J31, J45, J46, E26

**Acknowledgements:** Foremost, I would like to thank my advisor, Professor Raul Ramos for all his support and patience, and also thank my family for having trusted and supported me throughout my career. I thank all the people who supported me in this great challenge, since without their support the result would not have been possible.

# Contents

1. Introduction .....	3
2. Literature review .....	5
2.1. Theoretical.....	5
2.2. Empirical evidence .....	5
3. Data and methodology .....	6
3.1. Data sources .....	6
3.2. Methods.....	7
3.2.1. Oaxaca – Blinder (1973) .....	7
3.2.2. Melly (2006) .....	9
3.2.3. Firpo, Fortin, and Lemieux (2009) .....	10
3.3. Application .....	10
4. Results .....	12
4.1. Descriptive evidence .....	12
4.2. Decomposition analysis results.....	14
4.2.1 Decomposition analysis in the mean (Oaxaca – Blinder) .....	14
4.2.2. Decomposition analysis along the distribution .....	16
5. Conclusion.....	20
6. References .....	22
7. Appendix .....	25

## 1. Introduction

Despite the remarkable growth of the Peruvian economy in the last 20 years, informality continues to be a structural problem that is shared by developing countries. According to INEI (National Institute of Statistics and Informatics), 7 out of 10 individuals of working age are informal, a situation that does not guarantee basic aspects such as the social security system, health insurance, access to new markets through exports, access to credit, benefits tributaries, among others. This situation is accompanied by various peculiar characteristics of the Peruvian labor market; (Cling, et al., 2014) indicates that 75% of income comes from wages, and that more than half of urban workers work in companies in the informal sector. These data suggest that there is a high relationship between wages and informality. Furthermore, when in Peru around 99% of formal companies are micro and small enterprises, which employ 58% of the economically active population (PRODUCE, 2018), it is for this reason that the formalization of small companies is sought, since if achieved, it would generate an improvement in the labor market and high tax revenues. On the other hand, according to data from Peru, in the last year 2.1% and 2.2% of jobs were generated for men and women respectively; these jobs were especially for workers with non-university and university higher education (MTPE, 2020), this implies that there has been an increase in the education of workers, these being more qualified. Despite the high growth of developing economies, Peru being 4.8% in the last 10 years; countries in Latin America continue to have the highest rates of wage inequality (ILO, 2018; Wodon & Briere, 2018), that is, a negative relationship between them still persists. Among all the characteristics of the labor market, we will focus on analyzing wage gaps, more specifically in the formal and informal sectors.

Globally, great progress has been made in gender equality, which is why from the Global Gender Gap Index 2020, we can see that there is no country that has closed its gap completely (WEF, 2019), the first places are occupied by the Nordic countries and Nicaragua. Furthermore, Peru is in position 66 with an index of 0.714. It is also important to mention that Peru obtained 0.65 in the subindex of the area of economic participation and opportunities, which means that women are 35% less likely to have equal economic participation. Furthermore, this gender inequality generates a cost in national wealth (Lange et al., 2018) since human capital is measured as a cost of the future income of the labor force. It is important to mention that in Latin America and the Caribbean women receive an average labor income of 17% lower than men, having the same characteristics; this occurs despite significant long-term changes, since on average women with higher education exceed that of men, that is 40% of working women have higher education compared to 25% of men (ILO, 2019)

Currently in Peru there is a strong persistence of a wage gap between men and women over time, despite the fact that women have increased their level of labor participation by 73.3% and their level of education is now very high, this wage gap it has gone from 32.2% in 2004 to 28.6% in 2018 (MTPE, 2020), that is, on average, men have incomes around S/. 1589<sup>1</sup> and the women of S/. 1135. It is also important to consider the segmentation of the labor market, since having 72.4% informality, we can clearly see that the labor market is divided into two, on the one hand the informal sector, where labor regulations are non-existent and the characteristics of individuals are similar: low level of education, low income, no contract, among others; and on the other hand the formal sector where labor regulations protect workers against various risks. But in both there is wage discrimination, with a greater emphasis on the informal sector, since people with low educational levels and in a developing

---

<sup>1</sup> The exchange rate is around: \$ 1 = S / .344, that is to say \$ 462 for men and \$ 331 for women on average

economy such as Peru, it is common to see that men have a job, and women work at home. So we can distinguish that there is a double punishment in the informal sector for women; one for working informally, since they do not receive work benefits, so they earn very little money and a second punishment for being a woman, since in this sector there is more acceptance for men than for women because of physical abilities are more important than cognitive skills in this sector.

Unlike previous studies that focus on market structure, this research contributes to the little evidence that exists in labor market segmentation, which makes sense from dual labor market research (Doeringer, 1971) . In this study we are going to analyze the gender wage gap in 4 segments: public informal, public formal, private formal and private informal, for which we use various decomposition methods, on average and in along the wage distribution. We use data from the ENAHO (National Household Survey, 2018), which already generates a division of formal and informal individuals, a methodology carried out by the INEI and the ILO (International Labour Organization), the population we use is that of the economically active population, individuals between 14 and 65 years. The results obtained reveal that there is a significant wage gap between male and female workers in all segments, largely due to differences in the endowments of unobservable characteristics; except for the formal public segment where the gap is positive, that is, the wage differences are in favor of women, this is somewhat expected since in this sector the rules are applied as established by law, and since women have greater educational level, generates higher wages for them. Another important finding is that the level of occupation and sector are the variables that generate these salary differences, since as Peru is a very heterogeneous country, each sector and occupation have salaries already defined by the market, so these are variables that determine wages in the labor market. On the other hand, the results of the regression by quantiles show the effect called "glass ceiling" since the wage gap is wider at the top of the distribution and the "sticky floor" effect since the wage gap is wider in the lower part of the wage distribution, both effects for the formal public segment and only the "glass ceiling" effect for the informal public and private informal segments. In addition, for the private formal segment there is a high gap against women but stable (almost linear) along the distribution, this is due to informal practices in the public sector.

The rest of the master's thesis is structured as follows. In the literature section, we examine wage differences in the formal and informal segments, in addition to studies of the gender wage gap both in the mean and along the distribution. In the third section we present the data and the methodologies used, while in the fourth section we present the evidence obtained and end by summarizing the main conclusions.

## 2. Literature review

### 2.1. Theoretical

Wage discrimination occurs when two people with the same skills perform similar jobs, but are treated differently by the employer, this difference can have results that translate into different types of remuneration. However, wage gaps cannot only be attributed to discrimination. That is why since seminal work (Gary S. Becker, 1957) he mentioned two important aspects to explain wage differentials between workers: First the theory of human capital where she explains the differences in productivities between men and women; and second, female discrimination in the labor market, which causes inefficiencies. Later, more developed work is done, starting with the decomposition of Oaxaca Blinder (Oaxaca, 1973) that considers the Mincer equation for workers, estimating the differential in wages between men and women, decomposing the wage differential into one part explained and another unexplained, and finds that this differential in average wages is quite large. Subsequently, extensions were made to the Oaxaca-Blinder methodology, since the latter applies decomposition techniques to the mean. Therefore Machado & Mata (2005) use a method based on the estimation of marginal wage distributions consistent with a distribution estimated by regression by quantiles along wage distribution. Then Albrecht et al. (2009) extend the model making corrections for selection bias. Subsequently, estimators of unconditional covariate functions were proposed by Melly (2006) using Monte Carlo simulations that work well on finite samples, and also decompose the wage differential into an explained part (characteristics) and an unexplained part (coefficients). Finally Firpo et al. (2009), proposes to evaluate the changes in the quantiles of the unconditional distribution, for which it uses an influence function regression (RIF), they decompose the wage differential in the mean and along the distribution, also in a part explained and unexplained.

### 2.2. Empirical evidence

The concept of segment in the labor market was introduced by Peter B. Doeringer (1971.) Since their seminal contributions have related the gender gap to this concept dividing the labor market by formal and informal sector and by public and private sector. For example Cho & Cho (2011) show a substantial drop in size of the wage gaps in Korea, which is reflected by the unobserved capacity of the workers and considers that there is a simultaneity in the structure of the labor market and gender discrimination. Also, Ben Yahmed (2016) mentions that in Brazil the informal wage gap is 13%, while for the formal sector it is 5%, this is because women have better observable characteristics than men, such as education. Furthermore, Giordano et al. (2011) shows that the premium wage is higher for women and less developed countries in public sector workers; for Australia Barón & Cobb-Clark (2010) analyze the wage gap in the public and private sector, find an important effect in occupational segregation where the wage gap in low-wage workers is explained by the differences in the characteristics related to the wages but not fully explained by high-wage workers.

More recent research has also expanded the analysis beyond the mean considering gender gaps along the wage distribution and specific cases. For example, for the United Kingdom, Olsen & Walby (2004) show segregation in certain occupations in women and in small companies, which were responsible for 36% of the gender pay gap. In Greece, Livanos & Pouliakas (2012) calculate the gender wage gap considering education the most important factor, they show that this gap is around 8.4%, since women generally opt for a less risky education, that is why that women obtain lower wage premiums in the

labor market. Beaudry & Lewis (2014) consider the importance of people's abilities (cognitive, interpersonal and physical), it shows that the wage gap decreased during the 80s, while returns to education increased, that is, this decrease it was because people were able to adapt to technological change. Gauchat et al., (2012) and Oostendorp (2009) find that the gender wage gap tends to decrease with respect to globalization and trade, especially in richer countries. In Latin America these studies are more representative, since there is a much greater wage inequality, for example, in Chile, Montenegro (2001) uses the methodology of Machado & Mata (2005), it is shown that in the lower part of the salary distribution women have higher returns to education, but similar returns to men in experience; Furthermore, it is shown that at the bottom there is a gap of 10% and at the top around 40%. In Uruguay Borraz & Robano (2010) also uses the Machado and Mata methodology using the extension of Albrecht et al. (2009) shows that the gender wage gap increased throughout the distribution, showing the effect called "ceiling glass", it was also found that if all women worked, the wage gap could be greater.

In Peru there is little research that talks about the gender gap in the labor market. The most relevant research was carried out by Ñopo (2009) where it uses the matching comparison methodology and shows that one in four workers has characteristics that are not comparable to workers of the opposite sex, and that the wage gap is approximately 28%. Pozo (2017) uses the Machado and Mata methodology and finds that there is a decrease in the gap as we advance to the upper parts of the distribution.

In the literature we find that the wage gap can be studied taking into consideration different aspects, for example, by urban area, by occupational segregation, by type of skills, among others. In all cases we find a significant gap that shows that this problem must be studied in depth.

### **3. Data and methodology**

#### **3.1. Data sources**

The source from which the microdata is extracted in this study is the ENAHO (National Household Survey), the ENAHO is designed as a cross-sectional database, where we have quarterly and annual information, which is designed by the INEI (National Institute of Statistics and Informatics). The objective of this survey is to know the evolution of poverty and the living conditions of the households with coverage at the national level<sup>2</sup>, in the urban and rural areas and in the 24 departments of Peru. ENAHO is the main source of information for the generation of indicators such as the characteristics of housing, individuals, health, education, employment, among others. Something important is that the survey has maintained the same format since 2004, so the indicators can be compared year after year.

This survey is subdivided into sections, one of which is employment (Module 500), which we use, in this module the target population is individuals over 14 years of age, in our case we have only considered individuals between 14 and 65 years of age, the which represents the PEA (Economically Active Population). The most important variable to consider is informality, which is built by INEI, based on the methodology proposed by the ILO, this methodology takes into account the

---

<sup>2</sup> Members of the armed forces who live in barracks, camps, ships, and others are excluded from the survey. People residing in collective housing (hotels, hospitals, religious asylums and cloisters, prisons, etc.) are also excluded.

characteristics of the business or establishment where the individual works, and there is information from the year 2012. It is also made an update in the survey in relation to the sector where the individual works, which is based on the International Standard Industrial Classification of all economic activities (ISIC rev. 4); In addition, for the occupation of the individual, the International Standard Classification of Occupations (ISCO-2015) is considered, both cases are used to make them internationally comparable.

In addition, the data source provides information on the characteristics of the people: age, sex, experience, marital status and education; related to employment: occupation, type of contract and if the job is part-time full-time; related to geography: urban or rural, division of macro regions, and finally related to the company: sector where it operates, company size and whether the company is formal or informal. Wage information is used as a transformed variable to obtain hourly wages for the main occupation, which is calculated by dividing the monthly salary by the number of hours worked in a month. It's important mention that wages are imputed<sup>3</sup> and deflated<sup>4</sup>, it also includes overtime, bonuses, mobility, commissions, among other types of income. For the construction of the data, people older than 65 years were excluded and only individuals who have reported any income were considered. The final sample of the information was 26,806 observations, which correspond to 15768 men and 11038 women, also 15202 observations in the informal sector and 11064 observations in the formal sector, in addition to 19689 observations in the private sector and 6025 in the public sector. Descriptive statistics for the sample can be found in Table A.1 in the Appendix.

### 3.2. Methods

This research applies three different decomposition techniques to analyse the differential in wages between men and women. The first is the Oaxaca-Blinder methods (Blinder, 1973; Oaxaca, 1973), which allows detailing the decomposition into observable explanatory variables and unobservable characteristics between men and women in the mean. The second and third methodology are based on determining the wage differential along the distribution, in the first case we have applied Melly (2006) who uses unconditional covariate functions using Monte Carlo simulations and for the second case we have applied Firpo et al. (2009) that use a recentered influence function.

#### 3.2.1. Oaxaca – Blinder (1973)

The Oaxaca-Blinder methodology is based on separating the estimation in two, one in the difference existing from a common component (explained), and the other due to other factors (unexplained). We assume the mincerian equation. The expression can be expressed as follows:

$$w_i = X_i\beta + \varepsilon_i \quad (1)$$

---

3 The imputation process is carried out in two stages: a) First stage corresponds to the allocation of qualitative characteristics to household members who did not report the education (chap. 300), health (chap. 400) and employment (chap. 500) modules, leaving the quantitative variables as missing values. The Hot Deck technique is used to impute the qualitative variables. b) Second stage, values are assigned to the quantitative variables that were recorded as missing values (information not declared by the informant). For the imputation of the quantitative variables, the Median Matrices technique is used, thus guaranteeing that the assigned value is not influenced by extreme values.

4 Deflated is the process of transforming from nominal monetary values to real monetary values (at constant prices for a certain period), through the application of a price index that eliminates the effect of prices in the analysis period.

Where  $w_i$  represents the logarithm of the hourly wage of worker  $i$  and  $X_i$  represents the different individual characteristics, characteristics of employment and companies. First we separate the equation between men and women where  $E(w)$  is the expected value of the variable, to later take differences:

$$D = E(W_M) - E(W_F) = E(X_M\beta + \varepsilon_M) - E(X_F\beta + \varepsilon_F) \quad (2)$$

By property of the expected value we have the following:

$$E(X_i\beta) = E(X_i)\beta \quad \text{and} \quad E(\varepsilon_i) = 0 \quad (3)$$

Therefore, we have the following:

$$D = E(W_M) - E(W_F) = E(X_M)\beta - E(X_F)\beta \quad (4)$$

Instead of the expected values we will consider the average values:

$$D = E(W_M) - E(W_F) = (\bar{W}_M) - (\bar{W}_F) \quad (5)$$

$$\bar{W}_M = \bar{X}_M\beta_M ; \bar{W}_F = \bar{X}_F\beta_F \quad (6)$$

Replacing, we have the following:

$$D = \bar{W}_M - \bar{W}_F = \bar{X}_M\beta_M - \bar{X}_F\beta_F \quad (7)$$

Adding and subtracting  $\bar{X}_F\beta_M$  :

$$D = \bar{W}_M - \bar{W}_F = (\bar{X}_M - \bar{X}_F)\beta_M + \bar{X}_F(\beta_M - \beta_F) \quad (8)$$

In general, equation (1) is estimated using robust ordinary least squares for both the joint sample of men and women separately, and then decomposing it into the difference in average hourly wages, as in equation (8). Where the  $\bar{W}_M$  and  $\bar{W}_F$  are the average hourly wages of men and women. With this we obtain the decomposition, which is divided into two parts, the first  $\bar{X}_M$  and  $\bar{X}_F$  (left part) as the part explained by the differences between the observed variables and the second  $\beta_M$  and  $\beta_F$  (right part) the unexplained part which corresponds to the effect of the coefficients.

Both components can be decomposed, this in order to distinguish their contribution to each explained variable. This decomposition presents a problem, since when applying this procedure, the choice of a certain category for the dummy variables can affect the decomposition results, as shown in Oaxaca & Ransom (1994). For this reason, it is solved by applying the normalization of the dummy variables suggested by Yun (2005), this solves the problem and makes an adequate estimate of the real contribution of each variable to the differential.



### 3.2.2. Melly (2006)

The proposal made by Melly (2006) to analyze gender wage gap is based on a regression by quantiles, which was initially proposed by Koenker & Bassett (1978), which makes a more flexible analysis than estimation of OLS. Regression by quantiles estimates the effect of the explanatory variables in each percentile of the conditional distribution, this idea was developed by Machado & Mata (2005). Later Melly proposes a procedure to decompose the differences of the quantiles into an unconditional distribution. Melly's estimator uses Monte Carlo simulations, this estimator will be identical to the Machado and Mata estimator if the number of simulations goes to infinity, an advantage of Melly is its consistency is asymptotically normal. The procedure is the next:

First, we estimate a conditional distribution with a regression by quantiles:

$$Q_{\theta}(W/X) = X_i\beta(\varphi) \quad (9)$$

where  $\varphi \in (0,1)$  represents the number of quantiles.

Quantile regression is based on minimizing the sum of the absolute values of the deviations between the value of the salary and its predicted value. The quantile regression model assumes that the conditional quantile of  $w_i$  is linear in  $X$ . The vector of coefficients for  $\beta(\varphi)$  is estimated as follows:

$$\min_{\beta(\theta)} \left\{ \sum_{w_i \geq X_i\beta(\varphi)} \varphi |w_i - X_i\beta| + \sum_{w_i \leq X_i\beta(\varphi)} (1 - \varphi) |w_i - X_i\beta| \right\} \quad (10)$$

Second, the conditional distribution is integrated over the ranges of the covariates. Integrating on all quantiles and on all observations,  $\hat{\beta}(\hat{\beta}(\varphi_1), \dots, (\hat{\beta}(\varphi_j), \dots, (\hat{\beta}(\varphi_J)))$ , where  $j=1, \dots, J$ . Furthermore, the unconditional estimator  $\theta$  of the dependent variable is given by:

$$q(\theta, X, \beta) = \inf \left\{ q: \frac{1}{n} \sum_{i=1}^N \sum_{j=1}^J (\varphi_j - \varphi_{j-1}) (X_i \hat{\beta}(\varphi_j) \leq q) \geq \theta \right\} \quad (11)$$

Now we can estimate the counterfactual distribution and separate the difference of each quantile from the unconditional distribution in: part unexplained by the coefficients and part explained by characteristics:

$$q(\theta, X_M, \beta_M) - q(\theta, X_F, \beta_F) = [q(\theta, X_M, \beta_M) - q(\theta, X_M, \beta_F)] + [q(\theta, X_M, \beta_F) - q(\theta, X_F, \beta_F)] \quad (12)$$

Where the first bracket represents the effects of differences in coefficients (discrimination) and the second bracket represents the effects of differences in the distribution of the characteristics. In this case, as in the Oaxaca-Blinder methodology, the variables used for the estimation are the same, that is, characteristics of workers, characteristics of employment and companies.

### 3.2.3. Firpo, Fortin, and Lemieux (2009)

This methodology proposes an improvement to the development of the decompositions of the differences between the distributions of a variable. Which is to do an unconditional quantile regression using recentered influence function (RIF). The technique is developed in three steps, the first is to make a substitution of the dependent variable for a transformation, which is the recent influence function (RIF), second, we have to divide the salary distribution into a composition and structure effect using a weighted procedure (the weights are estimated), third, is to make the estimation using the Oaxaca Blinder procedure.

The influence function measures the effect on distribution statistics of small changes in the underlying distribution. Fortin (2011) proposes that the recent influence function should be  $RIF(W; \varphi_\theta) = \varphi_\theta + IF(W; \varphi_\theta)$ , they define unconditional quantile regression as the conditional expectation of  $RIF(W; \varphi_\theta)$  dado  $X: E[RIF(W; \varphi_\theta)|X]$ .

The influence function (IF) for a quantile  $\theta_t$  is equal to:

$$IF(W; \varphi_\theta) = \theta - l\{W \leq \varphi_\theta\}/f_y(\varphi_\theta) \quad (13)$$

Where  $f_y$  is the marginal density function of  $W$ . Since the recent influence function is  $RIF(W; \varphi_\theta) = \theta_t + IF(W; \varphi_\theta)$ , we can say the following:

$$RIF(W; \varphi_\theta) = \varphi_\theta + \theta - l\{W \leq \varphi_\theta\}/f_y(\varphi_\theta) \quad (14)$$

The estimated coefficients can be calculated using the Oaxaca Blinder decomposition methodology. Firpo et al. (2007) show that a Blinder-Oaxaca type decomposition based on RIF-regression estimates can be approximated for any distributional statistic, including quantiles. The decomposition is as follows:

$$D(\theta) = \bar{X}_M \beta_M - \bar{X}_F \beta_F \quad (15)$$

$$D(\theta) = (\bar{X}_M - \bar{X}_F) \beta^*(\theta) + \{\bar{X}_M(\beta_M(\theta) - \beta^*(\theta)) + \bar{X}_F(\beta^*(\theta) - \beta_F(\theta))\} \quad (16)$$

Where  $\beta_M(\theta)$  and  $\beta_F(\theta)$  are the unconditional quantile regression coefficients in men, women and in the group pool. The  $\bar{X}_M$  and  $\bar{X}_F$  are the average observed characteristics for men and women. The first component on the right side is caused by differences in characteristics, referring to the explained component, the second component refers to the effects of the coefficients, that is, the unexplained component. Interpretation of the results should be taken with care, especially with the unexplained component as it is subject to misinterpretation and different points of view.

### 3.3. Application

In this section we are going to explain the models that we are going to use in all the methodologies. Explanatory variables in empirical analysis include geographic characteristics, characteristics of individuals, as well as characteristics of work and establishment. The former serve as controls at the geographic level: zone (urban or rural) and macroregion (North West, Center, South West, South East, North East and Lima), controls for individuals: schooling (measured in years), marital status (single or

married), controls for work, such as the type of contract (indefinite, partial and without contract), time of employment (full time or part time), occupation (10 categories according to ISIC rev. 4), sector (6 categories according to ISCO-2015). Finally, the characteristics of the establishment, specifically for the type of company (small, medium, and large).

As previously mentioned, due to the heterogeneity of the Peruvian labor market, we have divided it into four segments: Public informal, private informal, public formal and private formal, each of these markets has different characteristics, which makes it necessary to carry out the analysis in a differentiated way.

### **Informal public**

In this first case we have 964 observations, it is the smallest segment in the labor market, where individuals work only in 3 sectors: mining, manufacturing and services; they also work in all occupations except the armed forces. It would be expected that in this sector the activities will be carried out as established by law, since we are in the public sector, but we find ourselves in formal employment that uses informal practices. We also see that the most performed occupations are administrative support, sales and elementary occupations. In this model we do not control for company size, since all observations are related to large companies.

$$\begin{aligned} \ln Wage_i = & \beta_0 + \beta_1 women + \beta_2 schooling + \beta_3 experience + \beta_4 experience^2 \\ & + \beta_5 married + \beta_6 urban + \beta_7 macroregion + \beta_8 contract \\ & + \beta_9 parttime + \beta_{10} sector + \beta_{11} occupation + \varepsilon_i \end{aligned} \quad (17)$$

### **Informal private**

In this case we have 12898 observations, and it is the largest segment in the labor market, since approximately 72% of informality is found in this segment, especially in small companies, which is a very common feature in developing countries. This segment is occupied by all sectors and all occupations except for the armed forces. This is the invisible activity of the State since it is the segment where taxes are not paid, for example: undeclared domestic work, ambulatory sales, among others. According to Perry et al. (2010), there are two causes for informality, one is via expulsion, that is, you cannot or do not want to incur greater expenses than being formal (bureaucracy for the formalization of a company), and the another is via exclusion, that is, the services provided do not compensate the payments made (lack of access to markets, financing, among others).

$$\begin{aligned} \ln Wage_i = & \beta_0 + \beta_1 women + \beta_2 schooling + \beta_3 experience + \beta_4 experience^2 \\ & + \beta_5 married + \beta_6 urban + \beta_7 macroregion + \beta_8 contract \\ & + \beta_9 parttime + \beta_{10} sector + \beta_{11} occupation + \beta_{12} Size\ of\ company \\ & + \varepsilon_i \end{aligned} \quad (18)$$

### **Formal Public**

In this case we have 4470 observations and it is the second smallest segment in the Peruvian labor market, it is the segment that is within the parameters regulated by the State, therefore it complies with all current regulations: tax, taxes, environment, labor, among others. This segment is occupied by all occupations, the most common jobs are: public officials (for example in ministries), teachers in public

schools, doctors in public hospitals, among others. Workers in this sector have medical insurance, in addition to the payment of pensions for social security. In this case we do not control by company size since it is assumed that the entire public system is considered as a large company. The specification is identical to the one shown in equation (17)

### **Formal Private**

In this case we have 6412 observations and the second largest segment of the Peruvian labor market. In general, there are very few small companies in this segment, as it is led by the large companies in Peru. For this reason it is found in all sectors and occupations except for the armed forces. Lima concentrates 50% of private formal employment in the country, and workers have access to social security, a contract, health insurance. All types of companies are concentrated in this segment, which is why we use all variables. All the data can be visualized in the appendix of table A.2. The specification is identical to the one shown in equation (18)

## **4. Results**

### **4.1. Descriptive evidence**

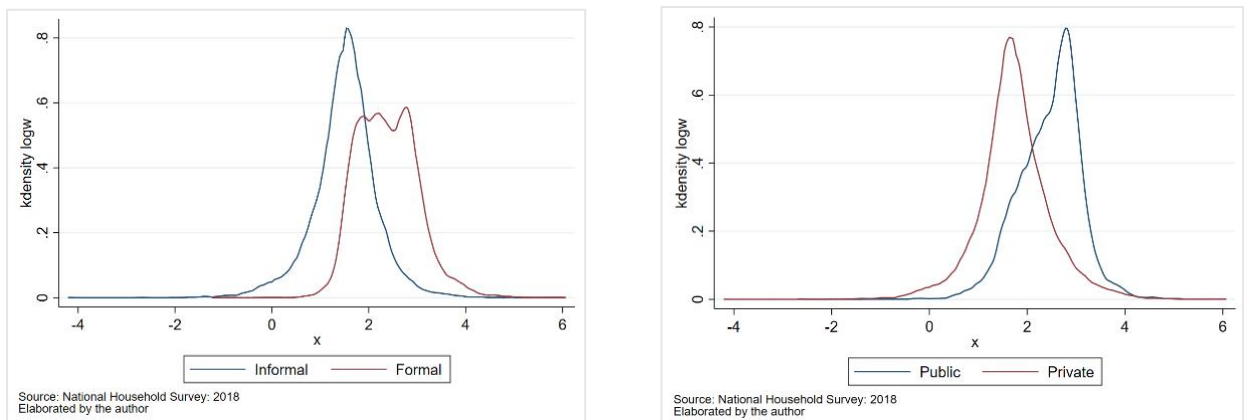
In this section we present statistical information on the main variables for men and women in the Peruvian labor market. The dependent variable is measured as the logarithm of the hourly wage, for men it is 2.0 on average and for women 1.8. In addition, the level of education for men is on average 12.5 years of study and for women it is 12.9 years, this implies that women are more educated than men, but receive lower wages, this may be due to discrimination. Another important fact is the years of experience that for men is an average of 18.5 years and for women it is 17.8 years. In the table it can be seen that the dominant sector is the informal sector, and it is located in the urban area, specifically in Lima, this shows that the labor market is centralized in its capital. We also see that the most relevant sector is services, in occupation, men have greater predominance in elementary occupations (Cleaners, drivers, kitchen helpers and others) and in the case of women there is greater predominance also in elementary occupations and also in the works of services and sales. It is important to mention that the majority of jobs are in small companies, which employ approximately 60% of the PEA (PRODUCE, 2018). We can see this information in the Table A.1.

Table A.2 presents the information disaggregated in the 4 segments to be analyzed, where it can be seen that the logarithm of the hourly wage occurs in the public formal sector, being 2.3 for men and women, and the lowest value is for the private informal sector, this being 1.3 for men and 1.2 for women; these two sectors are basically the opposite, since in one the formality guidelines are met at 100%, but in the other case informality is also shown at 100%; as mentioned before, informality represents around 72% of the labor market, which is why it is very important to analyze them separately. The other variables show the information in proportions in the 4 segments for both men and women.

In Figure 1 we show the kernel distributions in the division of the labor market structure (informal and formal) and (public and private) in the logarithm of hourly wages, we see in the graphs that there is no similar behavior that of the average shown in figure 2, this gives us an indication that the labor market should be analyzed separately, since they have independent behaviors. In the graph on the left (informal and formal) we see that the formal sector is skewed to the right, this means that there are higher wages in this sector, also the range of the lower limit is -1.23, which is less than that of the

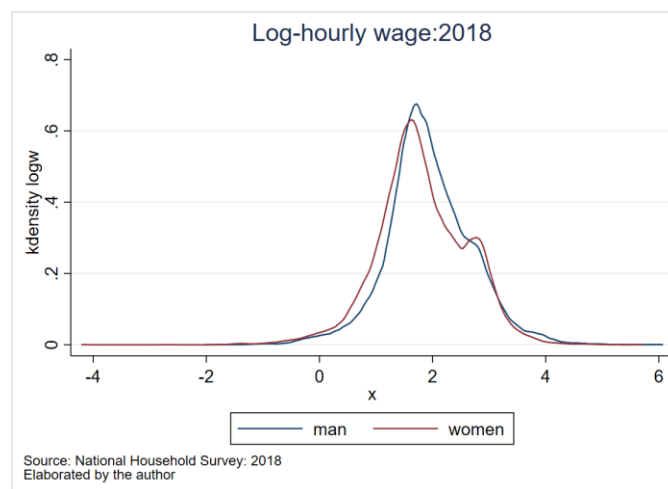
average in figure 2, this means that there are fewer people with low wages in this sector, quite the opposite occurs in the informal sector, where the lower limit range is -4.2, this means that most of the informal earn little money. In the graph on the right (public and private), we see that the public sector is more skewed to the right, since they on average have higher wages than those of the private sector, since most companies in the private sector are small companies and in the public sector are large companies. Regarding ranges, we see that in the public sector the lower limit is close to -2.0 and the upper limit is 4.0, which implies that many individuals have high wages, otherwise it happens in the private sector, where the range is much broader, because people are more distributed, in this case we see that there is an accumulation in the lower limit that indicates that individuals have very low wages.

Figure 1



Now in the In figure 2 we show the kernel density of the logarithm of hourly wages between men and women, we can see that they are very similar and slightly asymmetric to the right, with a mean for both cases close to 2.0 logarithmic points and the range fluctuates between -4.2 to 6, and we see that there is an accumulation on the left side of the distribution, this reflects there are many individuals with very low wages. We can also notice that on the average there are slightly more men than women and that they earn more, but we can also notice that in some parts of the distribution the opposite is the case.

Figure 2



## 4.2. Decomposition analysis results.

### 4.2.1 Decomposition analysis in the mean

#### Oaxaca – Blinder (1973)

Before starting to explain the results of the decompositions of each segment, we have made an estimate using OLS of all the joint data. The information related to the control variables is: education, married, experience, urban and macro regions. As can be seen in Table 1, the gender gap is much smaller in the restricted model, that is, in the model where only women are considered as explanatory (Column 1), but as we add more controls this gap is getting bigger, it varies from 12.8% to 17.7% in favor of men. In column 3, which is the most complete, we control it by informality, the negative sign reflects that the more informality, lower will be the hourly wage and the greater the gender wage gap against women, as in the private sector, since as I explained before this segment concentrates 72% of the labor market. It is important to mention that as the number of observations decreases, the gap increases, that is, this reduction affects women.

Table 1

Variables	(1)	(2)	(3)
Women	-0.128*** (0.00968)	-0.176*** (0.00804)	-0.177*** (0.00774)
Informal			-0.500*** (0.0213)
Private			-0.231*** (0.0115)
Informal*Private			0.0546** (0.0224)
Constant	1.941*** (0.00603)	0.220*** (0.02821)	1.284*** (0.0341)
Control variables	NO	YES	YES
Observations	26806	26172	25116
Men	15768	15436	15390
Women	11038	10736	9726
R-squared	0.007	0.358	0.442

Control variables are schooling, experience, experience^2, married, urban and Macro regions

Robust standard errors in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Elaborated by the author

In table 2 we show the results of the Oaxaca-Blinder decomposition of the 4 segments of the wage gap into the mean value of the wage distributions of men and women, without considering the sample selection bias. The information includes the average size of the wage differential and the values of the two components: characteristics (explained) and coefficients (unexplained). For the first case (informal + public), we show that the wage gap is around 28%, this being the largest of the 4 segments, that is, women in this segment have wages 28% less than men, the unexplained part it represents 76% (0.217) of the gap due to discrimination, and 24% (0.0692) is due to the explained part. In the second case (informal + private), it is the second segment with the largest wage gap and is the largest segment of the labor market since most are micro-companies, the unexplained part represents 101.6% (-2.10) and the part explained represents -1.56% (0.00324), despite the fact that the gap is less than the previous segment, the discrimination is very high due to the work carried out in this segment focusing more on

physical than cognitive abilities, that is, more for men than women. In the third case (Formal + public), it is the only segment where the wage gap is positive, that is, women have more wages than men and this is due to the high educational level of women, since occupations in this sector they are more for cognitive abilities than for physical abilities, the unexplained part represents -155%. (-0.0438) and the explained part 255% (0.0720), This is an expected result since in this segment the standards stipulated by "Ministerial Resolution N° 145-2019-TR" are met, through a methodological guide<sup>5</sup> developed by the MTPE (Ministry of Labor and Employment Promotion). In the fourth case (Formal + private), we show that the wage gap is 17.9%, this segment is the most complete in the labor market since there are all types of companies, although these are few, the unexplained part represents 115.4% (-2.06) and the explained part -15.4% (0.0275), we clearly see that it is the segment with the highest discrimination in the labor market. We conclude in this part that the private sector is the one with the most discrimination and only the formal public sector has a wage gap in favor of women.

Table 2

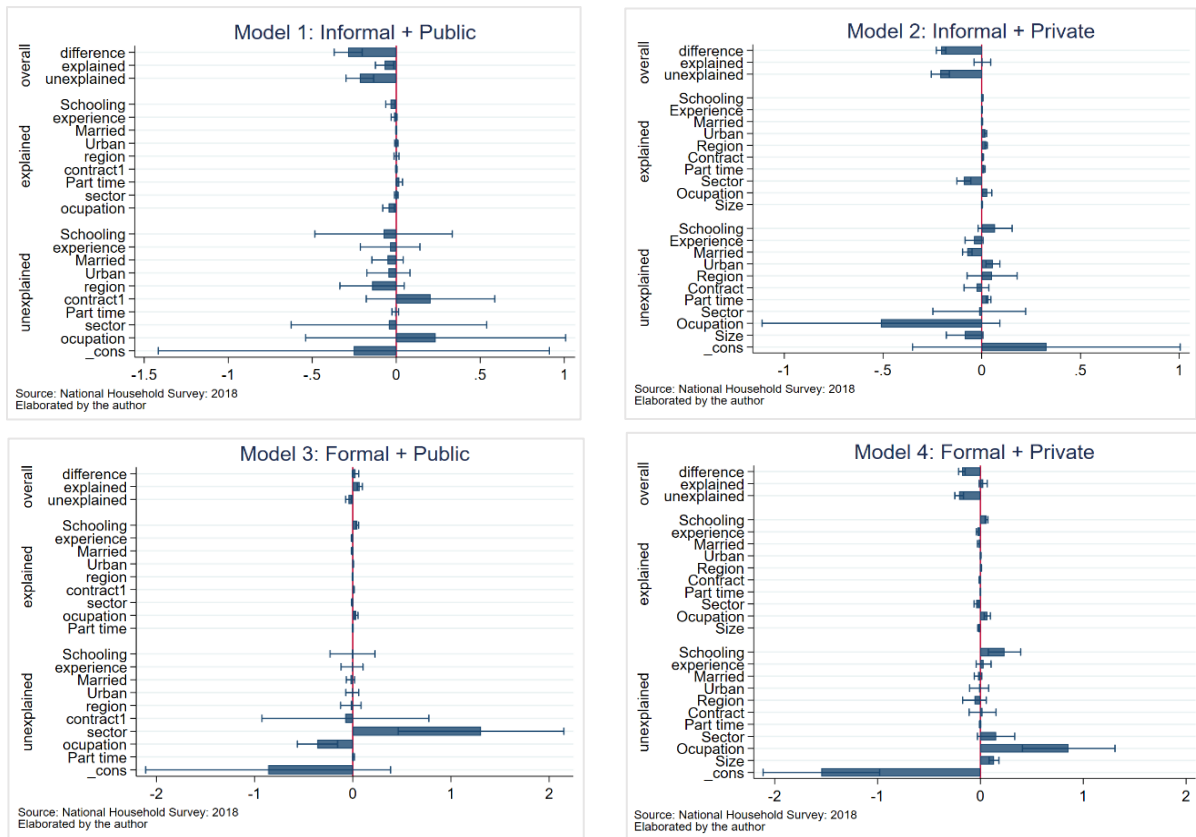
<b>Oaxaca - Blinder</b>				
<b>Variables</b>	<b>Informal + Public</b>	<b>Informal + Private</b>	<b>Formal + Public</b>	<b>Formal + Private</b>
Women	1.655*** (0.0291)	1.369*** (0.00985)	2.597*** (0.0105)	2.105*** (0.0129)
Men	1.941*** (0.0308)	1.575*** (0.00711)	2.569*** (0.0123)	2.284*** (0.0101)
Difference	<b>-0.286***</b> (0.0424)	<b>-0.206***</b> (0.0121)	<b>0.0282*</b> (0.0162)	<b>-0.179***</b> (0.0164)
Explained	-0.0692** (0.0280)	0.00324 (0.0215)	0.0720*** (0.0134)	0.0275 (0.0201)
Unexplained	-0.217*** (0.0422)	-0.210*** (0.0235)	-0.0438*** (0.0151)	-0.206*** (0.0217)
Observations	964	12,898	4,470	6,412
Standard errors in parentheses				
*** p<0.01, ** p<0.05, * p<0.1				
Elaborated by the author				

In figure 3 and in the appendix in table A.3 we see the detail of those mentioned in the previous table. In general, we see that the occupation and sector variables are that have the highest representativeness in wage gaps and especially in the unexplained part. For example, in the informal private segment, the most relevant variable is occupation in the unexplained part that has a value of -0.361 (see appendix table A.3), since in this sector physical abilities predominate over cognitive abilities, therefore, men have an easier getting a job with a good salary compared to women. In the case of the formal public segment where the gaps are positive, there are two relevant variables that make it positive: one is education (0.0467) in the explained part and the other is the sector (1,306) in the unexplained part, despite the fact that the coefficient is higher, the explaining part has a greater weight, therefore greater importance, which is why education in this sector is the variable that causes the gap to be positive in favor of women. The detail of the other interpretations can be seen in figure 3, where the longest bars

5 This guide is called " Methodological Guide for the Objective Valuation, without Gender Discrimination, of Jobs and Preparation of Tables of Categories and Functions "

are analyzed and also in the appendix of table A.3 where the value of the coefficients must be analyzed, but first you have to know which of the variables have greater weight in the salary difference.

Figure 3



#### 4.2.2. Decomposition analysis along the distribution

##### Melly (2006)

In table 3 we have the results obtained by performing the decomposition of men and women through a quantile regression which allows us to observe the variation in the distribution behind the analysis of means. The wage gap in this case is the difference between the logarithm of the hourly wage for men in a specific quantile of their distribution and the logarithm of the hourly wage for women in the same quantile of their distribution.

According to the table, the null hypothesis is clearly rejected for most quantiles and also throughout the quantile regression process. Therefore, we have estimated 100 separate quantile regressions in each sector through a bootstrap. Then, we have decomposed the differences between the quantiles of the unconditional distributions into a part explained by different characteristic distributions and an unexplained part by different coefficients (could be interpreted as discrimination).

As we can see in Table 3, we have that for the informal public, informal private and formal public the wage differential is negative and significant in all quantiles, that is, women earn less than men in all



scenarios, which is comparable with the results obtained in Oaxaca Blinder, which is reflected in our table, as the mean. But in the case of formal public, we have a positive sign up to the 60th percentile, then the negative sign appears, this tells us that at the top of the distribution, that is, individuals with very high wages, women earn less than men. This indicates that normally high-ranking positions in the public sector such as: Directors, Heads of public institutions, among others, are mostly held by men.

Table 3

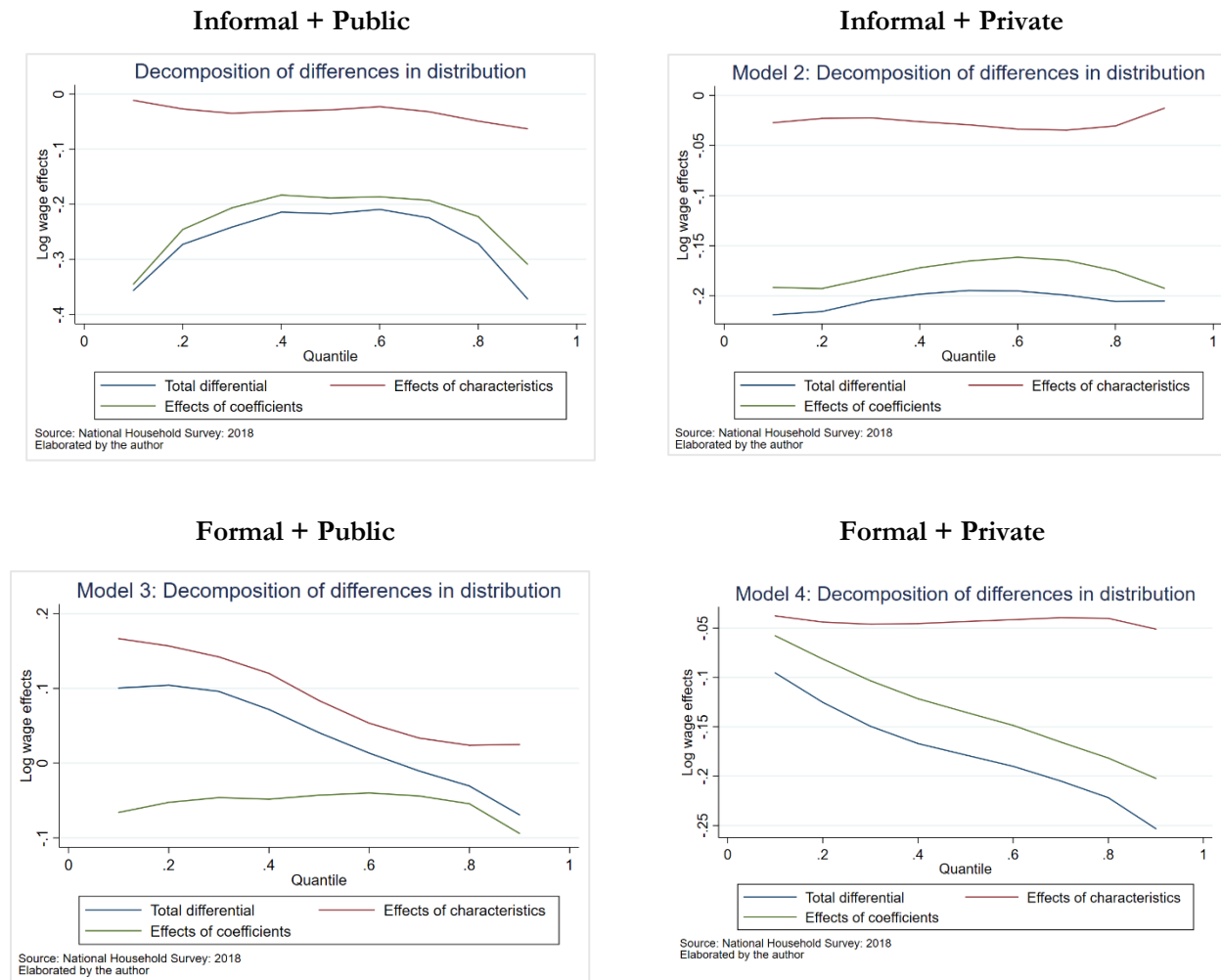
	Melly (Quantile regression)										Obs.
	Mean	10th	20th	30th	40th	50th	60th	70th	80th	90th	
<b>Informal + Public</b>											
Difference	-0.286***	-0.356***	-0.273***	-0.241***	-0.214***	-0.217***	-0.209***	-0.225***	-0.271***	-0.372***	964
<b>Informal + Private</b>											
Difference	-0.206***	-0.219***	-0.216***	-0.205***	-0.198***	-0.195***	-0.195***	-0.199***	-0.206***	-0.205***	12898
<b>Formal + Public</b>											
Difference	0.0282*	0.101***	0.104***	0.096***	0.072***	0.041***	0.013	-0.011	-0.031***	-0.069***	4470
<b>Formal + Private</b>											
Difference	-0.179***	-0.095***	-0.125***	-0.149***	-0.167***	-0.179***	-0.190***	-0.205***	-0.222***	-0.253***	6412

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Elaborated by the author

In figure 4 we have the detail of the explained, unexplained part and the total difference in each quantile (we can also see it in the annex to Table A.4). In addition, we mention that the graphs plot the decomposition results with a 95% confidence interval for all estimates. In the first case, informal public we see that the median (50th percentile) is very similar to the value obtained in the Oaxaca methodology, we also see an inverted “U” shape that indicates that in the lower and upper percentile the wage gaps they are larger, especially in the 10th (36%) and 90th (37%) percentiles, with this we demonstrate the effect called “glass ceiling” and “stick floor”, also according to the graph we see that the line of the coefficients is very similar to that of the differences, this indicates that there is a great effect because of unobservable characteristics (discrimination). In the second case, informal private, we also see that the median is similar to the value obtained by Oaxaca, in addition that there is not much fluctuation in the wage difference, this means that the gap is very similar in all the percentiles of the distribution which it varies between 19.5% and 21.6%, in addition to that the line of the effects of the coefficients is very similar to the total differential due to discrimination. This means that in this sector, which is the largest in the labor market, the gap is constant regardless of whether wages are high or low, this is because wages in this sector are close to the minimum wage, in addition there are more predominance of physical than cognitive abilities. In the third case, formal public, we also see that the median is similar to the value obtained in Oaxaca Blinder, the line related to the total differential is lower in the bottom and greater in the upper percentiles, this shows the effect called “ceiling of glass”, that is, when wages are low there is no discrimination, women earn 10% more than men, but while the percentiles increase, as in the 90th percentile, discrimination increases against women which can be seen in the line of the coefficients, so that women earn 6.9% less than men. In the fourth case, formal private, the median is exactly the same as that of Oaxaca Blinder, we also see that in this segment the effect called “glass ceiling” also occurs, since the greatest gaps are found in the highest percentiles, in this case we see that it is also due to the discrimination that exists in this sector, since the gap varies between 9.5% in the lowest percentile to 25.3% in the highest percentile.

Figure 4



It is important to mention that this analysis was also done by changing the dependent variable by the logarithm of average wages, and the results obtained are very similar (See appendix to table A.5). The analysis does not lead to two preliminary conclusions; On the one hand, the effects of wage differentials are strongly marked by discrimination in the four segments, especially in the upper percentiles, on the other hand, women, despite high economic growth, from the inclusion of international objectives against wage discrimination we see that they continue to face the so-called “glass ceiling” in the highest-ranking occupations.

**Dinardo, Fortin and Lemieux (2009)**

In table 4 we have the unconditional quantitative estimation using recentered influence function. This type of estimation can be done both in the mean and in the quantiles, as can be seen in the table of the results. In this case, the dependent variable is the logarithm of the hourly wage which is transformed to be replaced by the recentered influence function (Firpo et al 2009). We see that the results in the mean are exactly the same as in Oaxaca Blinder (see Table 2) and in relation to the values in the percentiles, these are very similar to those estimated in Melly (see Table N 3), the difference is that in this model, the quantiles at the bottom are greater than those estimated in Melly (2006) and the

quantiles at the top are less than those estimated in Melly, this means that, there is a greater wage gap in individuals with low wages, and a smaller gap in individuals with high wages. Furthermore, in most of the segments the coefficients are statistically significant at 95% in the distribution. An advantage that this model has, unlike Melly (2006), is that it shows the detail of the unconditional quantile decomposition of the effects of the characteristics and coefficients, that is, we can see which are the variables that have the greatest impact.

As mentioned above, the results are similar to those made in Melly (2006) when the total differential is analyzed. In this model we can note for example, in the first segment which is informal public, the greatest effect is in the explained part which is different from what is mentioned in the mean for the Oaxaca Blinder model, specifically in the pure explained characteristics (see annex in table A.5), one of the variables that most affects is education, therefore education would be the variable that generates a large wage gap in this segment, especially in the 10th and 90th percentiles, which they reflect 39% and 29% respectively of the wage gap. In the second segment, which is the informal private, it is clearly shown that the main cause of the high wage gap is caused by the unexplained part (discrimination), this is evident according to the international literature, since wages are already defined in this segments, where there is more preponderance for physical than cognitive abilities, is why there is a constant wage gap that ranges from 18.9% to 23%. For the third segment, which is formal private, we see that the gap continues to be positive until the 60th quantile and then becomes negative, this is evidence of what is called “glass ceiling”, the main factor for the wage gaps to be positive in favor of women is that in this segment there is a higher acceptance of cognitive than physical abilities, however, the influence of this factor is not homogeneous throughout the distribution, as can be seen in the 90th percentile. And finally, in the fourth segment, formal private, we see a smaller wage gap against women, unlike the other segments. In addition, in the first percentiles, the gap is very small, that is, there is very little difference between women's wages compared to men's, and we see that this is much more due to the unexplained part up to the 60th percentile, Later we see that the gap becomes much larger in the percentiles of the upper limit and this is mainly due to the explained part, emphasizing again on education, which has a greater effect on the characteristics of this segment.

Table 4

	Mean	10th	20th	30th	40th	Quantiles				90th
						50th	60th	70th	80th	
<b>Informal + Public</b>										
Total difference	-0.286*** (0.0417)	-0.392*** (0.0627)	-0.279*** (0.0503)	-0.215*** (0.0463)	-0.214*** (0.0441)	-0.229*** (0.0453)	-0.203*** (0.0495)	-0.234*** (0.0573)	-0.291*** (0.0613)	-0.291*** (0.0706)
Explained	-0.666*** (0.169)	-0.471*** (0.0800)	-0.384*** (0.102)	-0.383*** (0.142)	-0.472** (0.220)	-0.672** (0.307)	-0.833* (0.460)	-1.107*** (0.284)	-0.990*** (0.235)	-1.148*** (0.233)
Unexplained	0.380** (0.175)	0.0783 (0.0794)	0.105 (0.105)	0.168 (0.146)	0.258 (0.224)	0.443 (0.310)	0.630 (0.463)	0.873*** (0.288)	0.699*** (0.240)	0.857*** (0.240)
<b>Informal + Private</b>										
Total difference	-0.206*** (0.0121)	-0.220*** (0.0256)	-0.227*** (0.0185)	-0.196*** (0.0138)	-0.189*** (0.0121)	-0.189*** (0.0114)	-0.190*** (0.0112)	-0.191*** (0.0116)	-0.198*** (0.0139)	-0.237*** (0.0205)
Explained	-0.0994** (0.0392)	-0.0107 (0.0613)	-0.0541 (0.0423)	-0.0679** (0.0314)	-0.0633** (0.0281)	-0.0657** (0.0265)	-0.0638** (0.0282)	-0.0895** (0.0384)	-0.121** (0.0536)	-0.226*** (0.0783)
Unexplained	-0.107** (0.0435)	-0.209*** (0.0717)	-0.173*** (0.0492)	-0.128*** (0.0362)	-0.126*** (0.0320)	-0.123*** (0.0299)	-0.126*** (0.0317)	-0.101** (0.0424)	-0.0773 (0.0584)	-0.0112 (0.0850)
<b>Formal + Public</b>										
Total difference	0.0282* (0.0161)	0.0631** (0.0298)	0.129*** (0.0255)	0.113*** (0.0251)	0.0826*** (0.0235)	0.0503*** (0.0195)	0.0166 (0.0163)	-0.00328 (0.0160)	-0.0290 (0.0177)	-0.0636*** (0.0245)
Explained	0.0275 (0.0382)	0.0716*** (0.0230)	0.182 (0.118)	0.0810 (0.0721)	0.0470 (0.0499)	0.0215 (0.0326)	0.00292 (0.0231)	-0.0127 (0.0193)	-0.0393** (0.0187)	-0.0740*** (0.0246)
Unexplained	0.000664 (0.0424)	-0.00850 (0.0312)	-0.0531 (0.122)	0.0320 (0.0770)	0.0356 (0.0558)	0.0288 (0.0387)	0.0137 (0.0289)	0.00937 (0.0257)	0.0103 (0.0264)	0.0104 (0.0359)
<b>Formal + Private</b>										
Total difference	-0.179*** (0.0163)	-0.0806*** (0.0143)	-0.109*** (0.0144)	-0.139*** (0.0153)	-0.184*** (0.0166)	-0.207*** (0.0185)	-0.205*** (0.0211)	-0.235*** (0.0248)	-0.212*** (0.0311)	-0.198*** (0.0403)
Explained	-0.0689* (0.0372)	0.0159 (0.0332)	0.00755 (0.0289)	-0.00211 (0.0294)	-0.00908 (0.0337)	-0.0329 (0.0451)	-0.0800 (0.0564)	-0.125* (0.0665)	-0.283*** (0.0949)	-0.135*** (0.0412)
Unexplained	-0.110** (0.0441)	-0.0965** (0.0388)	-0.117*** (0.0349)	-0.136*** (0.0360)	-0.175*** (0.0410)	-0.174*** (0.0534)	-0.125* (0.0654)	-0.110 (0.0757)	0.0706 (0.105)	-0.0624 (0.0571)

Standard errors in parentheses

\*\*\* p&lt;0.01, \*\* p&lt;0.05, \* p&lt;0.1

Elaborated by the author

We found that the education variable is the main factor for a large wage gap both in the explained and unexplained part, this is consistent with Chi et al. (2011) for urban Chinese women, Galego & Pereira (2010) women in Portugal, Colella (2014) for women of the European Union in general, among others.

## 5. Conclusion

Peru is an interesting case, since like most developing countries, one of the most important problems is informality, with Peru being 72.4%, that is, 7 out of 10 workers are informal. This high informality causes the labor market to be divided into two large groups (formal and informal), since each group has different characteristics. This study analyzes the gender wages differences between men and women in the formal and informal sector, which in turn is subdivided into public and private, so we have 4 segments to analyze (informal private, informal public, formal public and formal private). We used three decomposition methods where the idea was to decompose the logarithm of hourly wages into an explained part attributable to the characteristics of the model and an unexplained part attributable to the coefficients, which is considered as the portion of the gap due to discrimination.

To estimate the models we use information from the ENAHO (National Household Survey) for 2018, and we apply the following decomposition methods: Oaxaca Blinder (1973) which estimates the mean wage differential, also the method proposed by Melly (2006), in this case estimates the wage differential in a quantile regression and finally the method of Firpo et al. (2009) which applies the recentered

Influence Function (RIF) in a regression in the mean and in quantiles for the wage distribution in all cases.

For the mean wage differentials we have that three segments present a negative wage gap against women and that only the formal public segment presents a positive wage gap and this is because in this segment cognitive abilities are more valued than physical, reason why the education is the variable that has more influence. The informal private segment is the largest segment of the labor market since it is encompassed by the majority of micro enterprises, this is the opposite of the formal public segment since here physical abilities are valued more than cognitive, so the unexplained part has a greater influence, that is, discrimination. The informal public segment is the one with the largest wage gap, this segment is basically made up of large companies, that is, this segment is where the highest wages are paid, the unexplained part has the greatest influence, so the factor most relevant for this gap to be so great, is discrimination.

Now, for the wage differences in each quantile we use two methodologies: Melly (2006) and Firpo, Fortin, and Lemieux (2009), one of the main differences between them is that the last one shows the detail of the variables that had the greatest influence on the explained and unexplained part. The results obtained for both models are very similar and show that the wage gap between men and women changes along the distribution, where it can be seen that in the first segment (informal public), it is where the greatest gap occurs, we see that the wage differential along the distribution has an inverted “U” shape, this indicates that in this segment in the lower and upper quantiles the gap is much greater, showing the effects “glass ceiling” and “stick floor”, in addition to this gap being given in greater magnitude by the unexplained part, that is, discrimination. On the other hand, in the formal public segment, we mentioned before that in the average there is a positive gap, but in the analysis in each quantile we see that there is a positive gap up to the 60th quantile, strongly positioned by the explained part, especially due to the high level educational of women, later this gap becomes negative and this is again due to the unexplained part, ie discrimination. In addition, we found a strong glass ceiling effect in the formal public and private informal segments, that is, a very large gap in the upper percentiles (very high wages), this clearly reflects that despite the country's high level of growth and improvement in inequality, there is still discrimination against women, especially in the occupation of some important job that entails a high salary despite the fact that women have increased their educational level compared to men.

In general, we can say that in Peru, education has a strong impact on decisions about participation and wages in the labor market for men and women, and it continues to be the main reason for high wage inequality. It is important to mention that the results obtained are of high importance in the design of policies, since we identified in which segments women have lower wages and what are the reasons.

Future research, which analyzes the gender wage gap in a segmented labor market, should try to correct the selection bias for a better approach. Despite this limitation, we believe that the results presented are robust enough to be able to show evidence about wage gaps in the different segments. Furthermore, it is recommended that policy makers take this research into account and propose measures to reduce the gender wage gaps.

## 6. References

- Albrecht, J., van Vuuren, A., & Vroman, S. (2009). Counterfactual distributions with sample selection adjustments: Econometric theory and an application to the Netherlands. *Labour Economics*, 16(4), 383–396. <https://doi.org/10.1016/j.labeco.2009.01.002>
- Barón, J. D., & Deborah A., C.-C. (2010). Occupational Segregation and the Gender Wage Gap in Private- and Public-Sector Employment: A Distributional Analysis. *Economic Record*, 86(273), 227–246. <https://doi.org/10.1111/j.1475-4932.2009.00600.x>
- Beaudry, P., & Lewis, E. (2014). Do male-female wage differentials reflect differences in the return to skill? Cross-city evidence from 1980-2000. *American Economic Journal: Applied Economics*, 6(2), 178–194. <https://doi.org/10.1257/app.6.2.178>
- Ben Yahmed, S. (2016). Formal but Less Equal. Gender Wage Gaps in Formal and Informal Jobs in Urban Brazil. *World Development*, 101(16), 73–87. <https://doi.org/10.1016/j.worlddev.2017.08.012>
- Blinder, A. S. (1973). Wage Discrimination: Reduced Form and Structural Estimates. *The Journal of Human Resources*, 8(4), 436–455.
- Borraz, F., & Robano, C. (2010). BRECHA SALARIAL EN URUGUAY. *Revista de Análisis Económico*, 25, 49–77.
- Chi, W., Li, B., & Yu, Q. (2011). Decomposition of the increase in earnings inequality in urban China: A distributional approach. *China Economic Review*, 22(3), 299–312. <https://doi.org/10.1016/j.chieco.2011.03.002>
- Cho, J., & Cho, D. (2011). Gender difference of the informal sector wage gap: A longitudinal analysis for the Korean labor market. *Journal of the Asia Pacific Economy*, 16(4), 612–629. <https://doi.org/10.1080/13547860.2011.621363>
- Colella, F. (2014). Women's part-time–full-time wage differentials in Europe: an endogenous switching model. *MPR Paper No. 56735, RePEc, Munich*, 55287.
- Doeringer, P. B., & Piore, M. J. (n.d.). *Internal Labor Markets and Manpower Analysis*. Retrieved May 31, 2020, from [https://books.google.es/books/about/Internal\\_Labor\\_Markets\\_and\\_Manpower\\_Anal.html?id=a8s5YyWkaCwC&redir\\_esc=y](https://books.google.es/books/about/Internal_Labor_Markets_and_Manpower_Anal.html?id=a8s5YyWkaCwC&redir_esc=y)
- Firpo, S., Fortin, N. M., & Thomas, L. (2009). Unconditional Quantile Regressions. *Econometrica*, 77(3), 953–973. <https://doi.org/10.3982/ecta6822>
- Galego, A., & Pereira, J. (2010). Evidence on gender wage discrimination in Portugal: Parametric and semi-parametric approaches. *Review of Income and Wealth*, 56(4), 651–666. <https://doi.org/10.1111/j.1475-4991.2010.00413.x>
- Gary S. Becker. (n.d.). *The Economics of Discrimination*. Retrieved May 31, 2020, from <https://press.uchicago.edu/ucp/books/book/chicago/E/bo22415931.html>
- Gauchat, G., Kelly, M., & Wallace, M. (2012). Occupational Gender Segregation, Globalization, and Gender Earnings Inequality in U.S. Metropolitan Areas. *Gender and Society*, 26(5), 718–747. <https://doi.org/10.1177/0891243212453647>
- Giordano, R., Depalo, D., Coutinho Pereira, M., Eugène, B., Papapetrou, E., Perez, J. J., Reiss, L., & Roter, M. (2011). The public sector pay gap in a selection of Euro area countries. *ECB Working Paper Series*, 1406.

- ILO, (International Labour Organization). (2018). *Global wage report 2018/19: What lies behind gender pay gaps*. [https://www.ilo.org/global/publications/books/WCMS\\_650553/lang--zh/index.htm](https://www.ilo.org/global/publications/books/WCMS_650553/lang--zh/index.htm)
- ILO, (International Labour Organization). (2019). Las mujeres en el mundo del trabajo, ciudad de Córdoba, 1904-1919. *Labor Thematic Panorama*, 17(1), 1–204.
- Koenker, R., & Bassett, G. (1978). Regression Quantiles. *Econometrica*, 46(1), 33. <https://doi.org/10.2307/1913643>
- Lange, G., Wodon, Q., & Carey, K. (2018). The Changing Wealth of Nations 2018: Building a Sustainable Future. In *The Changing Wealth of Nations 2018: Building a Sustainable Future*. <https://doi.org/10.1596/978-1-4648-1046-6>
- Livanos, I., & Pouliakas, K. (2012). Educational segregation and the gender wage gap in Greece. *Journal of Economic Studies*, 39(5), 554–575. <https://doi.org/10.1108/01443581211259473>
- Machado, J. A. F., & Mata, J. (2005). Counterfactual decomposition of changes in wage distributions using quantile regression. *Journal of Applied Econometrics*, 20(4), 445–465. <https://doi.org/10.1002/jae.788>
- Melly, B. (2006). Estimation of counterfactual distributions using quantile regression. In *Review of Labor Economics* (Vol. 68, Issue 4). <https://doi.org/10.1002/jae.788>
- Montenegro, C. (2001). Wage Distribution in Chile: Does Gender Matter? A Quantile Regression Approach GENDER AND DEVELOPMENT Working Paper Series No. 20 Wage Distribution in Chile: Does Gender Matter? A Quantile Regression Approach. *Policy Research Report on Gender and Development*, 20, 1–35. <http://documents.worldbank.org/curated/en/600701468314070337/pdf/341330Gender0wp20.pdf>
- MTPE, (Ministerio del Trabajo y Promoción del Empleo). (2020). *Panorama Laboral 2014*. 117 p. <https://doi.org/10.1007/s13398-014-0173-7.2>
- Ñopo, H. (2009). The Gender Wage Gap in Peru 1986-2000: Evidence from a Matching Comparisons Approach. *Inter-American Development Bank, March*. <https://doi.org/10.2139/ssrn.1821913>
- Oaxaca. (1973). Male-Female Wage Differential in Urban Labor Markets. *International Economic Review*, 14(3), 693–709.
- Oaxaca, R. L., & Ransom, M. R. (1994). On discrimination and the decomposition of wage differentials. *Journal of Econometrics*, 61(1), 5–21. [https://doi.org/10.1016/0304-4076\(94\)90074-4](https://doi.org/10.1016/0304-4076(94)90074-4)
- Olsen, W., & Walby, S. (2004). Modelling Gender Pay Gaps. EOC Working Paper Series No. 53. *EOC Working Paper Series No. 17*, 17.
- Oostendorp, R. H. (2009). Globalization and the gender wage gap. *World Bank Economic Review*, 23(1), 141–161. <https://doi.org/10.1093/wber/lhn022>
- Perry, G. E., Maloney, W. F., Arias, O. S., Fajnzylber, P., & Saavedra-chanduvi, A. D. M. J. (2010). Informality: Exit and Exclusion. In *World* (Vol. 57, Issue c). <https://doi.org/10.1596/978-0-8213-7092-6>
- Pozo, J. M. Del. (2017). Has the Gender Wage Gap been reduced during the Peruvian Growth Miracle? A distributional Approach. *Documento De Trabajo (PUCP)*, 442.
- PRODUCE, (Ministerio de la Producción). (2018). *Las Mipyme en cifras 2017*. Recuperado de [Consulta 03 de mayo del 2020]. 228. <http://ogeie.produce.gob.pe/index.php/shortcode/oee-documentos-publicaciones/publicaciones-anauales/item/829-las-mipyme-en-cifras-2017>
- WEF, (World Economic Forum). (2019). *Global Gender Gap Report 2020: Insight Report*. <https://doi.org/10.1002/9781119085621.wbef350>

- Wodon, Q., & Briere, B. de. (2018). The cost of Gender Inequality “Unrealized Potential: The High Cost of Gender Inequality in Earnings.” *World Bank*, May. <https://doi.org/10.1596/29865>
- Yun, M. S. (2005). A simple solution to the identification problem in detailed wage decompositions. *Economic Inquiry*, 43(4), 766–772. <https://doi.org/10.1093/ei/cbi053>



## 7. Appendix

### 7.1. Description of variables and data own elaboration

Variable	Description	Data source
<b>Year</b>	It is the year when the survey was made	2018
<b>Ubigeo</b>	Represents the geographical location code	ENAHO
<b>Schooling</b>	Year of education of a worker	years in college, university, and not university (institute) are considered
<b>Age</b>	Year of worker	For workers over 14 years old
<b>Experience</b>	Knowledge or skills obtained throughout life	Age - Scholling - 5
<b>Ln wage</b>	Represents the logarithm of hourly wages	It is a transformed variable that is only used for occupation main, which is calculated by dividing the monthly wages by the number of hours worked in a month, wages are imputed and deflated, also includes wages for: overtime, bonus, mobility, commissions, among others
<b>Married</b>	Includes workers who are sharing your life with someone	1= married, 0= single
<b>Informal</b>	It includes workers who work in a company that does not comply with basic aspects such as: social security, health insurance, among others.	1= informal, 0= formal
<b>Part time</b>	Workers who work for a maximum of 20 hours a week are considered.	1= part time, 0=full time
<b>Private</b>	They are the companies that seek their profit in their activity and are not controlled by the State.	1= private sector, 0=public sector
<b>Urban</b>	It is considered an urban area when in a geographic space there are more than 2000 inhabitants	1=urban, 0=rural
<b>Macroregions</b>	It is an area that covers several regions with characteristics in common	1= Nor Oeste, 2= Centro, 3= Sur Oeste, 4= Sur Este, 5= Nor Este, 6= Lima
<b>Contract</b>	Lets know the type of contract of a worker	1= Indefined, 2= Partial, 3= Without contract
<b>Sector</b>	Represents the main occupation sector that the worker performed according to ISIC revision 4	1= Agriculture, forestry and fishing, =2 Mining and quarrying, 3= Manufacturing, 4= Construction, 5= Commerce, 6= Services
<b>Occupation</b>	Represents the main occupation that the worker performed according to ISCO 2015	1= Armed Forces Occupations, 2= Managers, 3= Professionals, 4= Technicians and Associate Professionals, 5= Clerical Support Workers, 6= Services and Sales Workers, 7= Skilled Agricultural, Forestry and Fishery Workers, 8= Craft and Related Trades Workers, 9= Plant and Machine Operators and Assemblers, 10= Elementary Occupations
<b>Size of company</b>	Represents the size of the company	1= Micro enterprises, 2= Medium, 3= Large

Elaborated by the author

## 7.2. Statistics

Table A.1. Descriptive analysis of the mean variables: 2018

Variables	Male				Female			
	Mean	Std. Dev.	Min	Max	Mean	Std. Dev.	Min	Max
Ln wage	2.0	0.7	-2.0	6.1	1.8	0.8	-4.2	5.7
Schooling	12.5	3.6	0	19	12.9	3.9	0	19
Age	36.0	12.5	14	65	35.8	12.0	14	65
Experience	18.5	13.1	0	60	17.8	12.9	0	60
Informal	0.7	0.5	0	1	0.7	0.4	0	1
Private	0.8	0.4	0	1	0.8	0.4	0	1
Married	0.5	0.5	0	1	0.5	0.5	0	1
Urban	0.8	0.4	0	1	0.8	0.4	0	1
Part time	0.2	0.4	0	1	0.4	0.5	0	1
<b>Macro Region</b>								
Nor Oeste	0.2	0.4	0	1	0.2	0.4	0	1
Centro	0.1	0.3	0	1	0.1	0.3	0	1
Sur Oeste	0.1	0.3	0	1	0.1	0.3	0	1
Sur Este	0.1	0.3	0	1	0.1	0.3	0	1
Nor Este	0.0	0.2	0	1	0.0	0.1	0	1
Lima	0.4	0.5	0	1	0.5	0.5	0	1
<b>Contract</b>								
Undefined	0.1	0.4	0	1	0.2	0.4	0	1
Partial	0.4	0.5	0	1	0.4	0.5	0	1
Without contract	0.5	0.5	0	1	0.5	0.5	0	1
<b>Sector</b>								
Agriculture, forestry and fishing	0.1	0.4	0	1	0.1	0.3	0	1
Mining and quarrying	0.0	0.2	0	1	0.0	0.1	0	1
Manufacturing	0.1	0.3	0	1	0.1	0.3	0	1
Construction	0.1	0.4	0	1	0.0	0.1	0	1
Commerce	0.1	0.3	0	1	0.1	0.4	0	1
Services	0.4	0.5	0	1	0.7	0.5	0	1
<b>Occupation</b>								
Armed Forces Occupations	0.0	0.1	0	1	0.0	0.1	0	1
Managers	0.0	0.1	0	1	0.0	0.1	0	1
Professionals	0.1	0.3	0	1	0.2	0.4	0	1
Technicians and Associate Professionals	0.1	0.3	0	1	0.1	0.3	0	1
Clerical Support Workers	0.1	0.3	0	1	0.1	0.3	0	1
Services and Sales Workers	0.1	0.3	0	1	0.3	0.4	0	1
Skilled Agricultural, Forestry and Fishery <sup>1</sup>	0.0	0.1	0	1	0.0	0.0	0	1
Craft and Related Trades Workers	0.1	0.3	0	1	0.0	0.2	0	1
Plant and Machine Operators and Assem	0.1	0.3	0	1	0.0	0.1	0	1
Elementary Occupations	0.4	0.5	0	1	0.3	0.5	0	1
<b>Size of the company</b>								
Mype	0.5	0.5	0	1	0.5	0.5	0	1
Medium	0.1	0.3	0	1	0.1	0.3	0	1
Large	0.4	0.5	0	1	0.4	0.5	0	1

Note: Sample includes individuals between 14 and 65 years. Observation weighted by expansion factor.

Ln wage represents the average hourly wage

Elaborated by the author based on INEI-National Household Survey

Table A.2. Distribution of the men and women of the main variables in means and proportions: 2018

Variables	Informal + Public		Informal + Private		Formal + Public		Formal + Private	
	Males	Females	Males	Females	Males	Females	Males	Females
Ln wage	1.6	1.5	1.3	1.2	2.3	2.3	2.1	2.0
Schooling	13.9	13.2	11.1	11.5	15.7	16.4	13.5	14.2
Age	35.6	32.9	33.0	32.3	45.6	44.2	37.3	34.8
Experience	16.7	14.7	16.9	15.8	24.9	22.8	18.9	15.5
Urban	42.0	39.4	50.7	31.6	49.3	45.1	62.7	34.3
Part time	4.5	10.6	12.1	9.4	5.6	6.6	4.0	3.1
<b>Type of contract</b>								
Undefined			0.1	0.0	24.53	26.9	18.0	8.7
Partial	49.6	48.8	10.0	7.5	25.15	23.4	44.7	25.0
Without contract	0.5	1.1	54.0	28.4	0.03	0.02	2.5	1.3
<b>Sector</b>								
Agriculture, forestry and fishing			15.1	5.1			4.8	1.8
Mining and quarrying	0.3	0.0	1.4	0.2	0.05	0	4.1	0.2
Manufacturing	0.5	1.3	8.2	3.9	1.08	0.27	13.8	5.5
Construction			14.7	0.3			6.0	0.9
Commerce			9.0	7.9			10.1	7.0
Services	49.1	50.1	15.8	18.6	51.8	46.8	26.4	19.5
<b>Occupation</b>								
Armed Forces Occupations					7.8	1.4		
Managers	0.4	1.7	0.02	0.03	2.1	1.0	0.6	0.2
Professionals	6.9	9.0	2.3	1.9	20.3	26.3	6.7	6.0
Technicians and Associate Professionals	4.4	5.5	2.1	3.8	3.8	5.1	12.1	5.6
Clerical Support Workers	10.2	6.0	2.6	2.0	4.4	5.0	7.5	7.6
Services and Sales Workers	12.8	9.4	14.2	6.6	6.0	5.1	9.7	9.1
Skilled Agricultural, Forestry and Fishery Workers	0.0	0.1	0.0	0.9	0.0	0.0	0.6	0.0
Craft and Related Trades Workers	0.0	0.7	2.2	10.7	0.5	0.0	6.8	0.9
Plant and Machine Operators and Assemblers	0.0	2.7	0.5	4.7	1.9	0.0	9.0	0.5
Elementary Occupations	15.2	15.0	12.0	33.6	6.2	3.1	12.2	5.0
<b>Size of the company</b>								
Mye			53.3	30.1			13.1	9.1
Medium			6.6	3.3			15.9	8.4
Large	50.1	49.9	4.2	2.5	49.7	50.3	36.1	17.4
Observations	439	525	8266	4632	2146	2324	4231	2181

Note: Sample includes individuals between 14 and 65 years. Observation weighted by expansion factor. The variables Ln wage, schooling, age and experience are in means, the others are in proportions. Ln wage represents the average hourly wage.

Elaborated by the author based on INEI-National Household Survey

### 7.3. Estimations

Table A.3. Oaxaca – Blinder

Oaxaca - Blinder								
Variables	Informal + Public		Informal + Private		Formal + Public		Formal + Private	
	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
<b>Overall</b>								
group_1	1.655***	(0.0291)	1.369***	(0.00985)	2.597***	(0.0105)	2.105***	(0.0129)
group_2	1.941***	(0.0308)	1.575***	(0.00711)	2.569***	(0.0123)	2.284***	(0.0101)
difference	-0.286***	(0.0424)	-0.206***	(0.0121)	0.0282*	(0.0162)	-0.179***	(0.0164)
explained	-0.0692**	(0.0280)	0.00324	(0.0215)	0.0720***	(0.0134)	0.0275	(0.0201)
unexplained	-0.217***	(0.0422)	-0.210***	(0.0235)	-0.0438***	(0.0151)	-0.206***	(0.0217)
<b>Explained</b>								
Schooling	-0.0343**	(0.0140)	0.00307*	(0.00178)	0.0467***	(0.00634)	0.0591***	(0.00696)
Experience	-0.0127	(0.00894)	0.000548	(0.000848)	-0.00789**	(0.00307)	-0.0290***	(0.00450)
Married	-2.51e-05	(0.000951)	0.00147	(0.00151)	-0.00689**	(0.00333)	-0.0162***	(0.00562)
Urban	0.00113	(0.00500)	0.0186***	(0.00286)	0.00462***	(0.00166)	0.00313*	(0.00181)
Macro regions	0.00188	(0.00764)	0.0226***	(0.00287)	-0.00229	(0.00149)	0.00583**	(0.00254)
Contract	0.000796	(0.00212)	0.00570***	(0.00155)	0.00859***	(0.00306)	-0.00620**	(0.00306)
Part time	0.0188*	(0.00964)	0.0133***	(0.00209)	-0.000340	(0.000878)	0.00186*	(0.000977)
Sector	0.000360	(0.00556)	-0.0903***	(0.0180)	-0.00654**	(0.00281)	-0.0403***	(0.0106)
Occupation	-0.0451**	(0.0178)	0.0272**	(0.0122)	0.0361***	(0.00908)	0.0677***	(0.0152)
Size of company			0.000989	(0.00121)			-0.0184***	(0.00338)
<b>Unexplained</b>								
Schooling	-0.0755	(0.209)	0.0677	(0.0440)	-0.00236	(0.116)	0.235***	(0.0805)
Experience	-0.0362	(0.0907)	-0.0389*	(0.0233)	-0.00780	(0.0571)	0.0318	(0.0369)
Married	-0.0518	(0.0474)	-0.0729***	(0.0122)	-0.0235	(0.0222)	-0.0220	(0.0186)
Urban	-0.0465	(0.0657)	0.0566***	(0.0179)	-0.00559	(0.0333)	-0.0118	(0.0474)
Region	-0.144	(0.0974)	0.0525	(0.0647)	-0.0198	(0.0528)	-0.0572	(0.0588)
Contract	0.204	(0.195)	-0.0272	(0.0320)	-0.0758	(0.434)	0.0211	(0.0666)
Part time	-0.00592	(0.0101)	0.0349***	(0.00574)	0.00804*	(0.00437)	-0.00219	(0.00334)
Sector	-0.0437	(0.297)	-0.0128	(0.120)	1.306***	(0.430)	0.153*	(0.0922)
Occupation	0.235	(0.395)	-0.510*	(0.307)	-0.361***	(0.105)	0.858***	(0.230)
Size of company			-0.0864*	(0.0473)			0.133***	(0.0244)
Constant	-0.252	(0.594)	0.327	(0.346)	-0.862	(0.636)	-1.545***	(0.289)
Observations	964		12,898		4,470		6,412	

There are some variables that are grouped, for example: Experience, Macro regions, Contract, Sector, Occupation and Size of company. Each variable with its respective dummies.

Standard errors in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Table A.4. Estimation using Melly (2006) using as dependent variables the logarithm of hourly wages

	Quantile regression								
	10th	20th	30th	40th	50th	60th	70th	80th	90th
<b>Informal + Public</b>									
Raw difference	-0.356*** (0.050)	-0.273*** (0.036)	-0.241*** (0.034)	-0.214*** (0.030)	-0.217*** (0.031)	-0.209*** (0.031)	-0.225*** (0.033)	-0.271*** (0.036)	-0.372*** (0.053)
Characteristics	-0.011 (0.049)	-0.027 (0.035)	-0.035 (0.037)	-0.031 (0.042)	-0.029 (0.046)	-0.023 (0.049)	-0.032 (0.050)	-0.049 (0.048)	-0.063 (0.060)
Coefficients	-0.345*** (0.041)	-0.246*** (0.030)	-0.206*** (0.028)	-0.183*** (0.031)	-0.188*** (0.035)	-0.186*** (0.038)	-0.193*** (0.044)	-0.222*** (0.052)	-0.309*** (0.063)
<b>Informal + Private</b>									
Raw difference	-0.219*** (0.020)	-0.216*** (0.013)	-0.205*** (0.011)	-0.198*** (0.010)	-0.195*** (0.008)	-0.195*** (0.008)	-0.199*** (0.009)	-0.206*** (0.011)	-0.205*** (0.018)
Characteristics	-0.027* (0.019)	-0.023*** (0.013)	-0.022*** (0.011)	-0.026*** (0.011)	-0.029 (0.011)	-0.034*** (0.012)	-0.035*** (0.013)	-0.031*** (0.014)	-0.013 (0.018)
Coefficients	-0.192*** (0.014)	-0.193*** (0.009)	-0.182*** (0.007)	-0.172*** (0.006)	-0.165*** (0.006)	-0.162*** (0.006)	-0.165*** (0.007)	-0.175*** (0.008)	-0.192*** (0.012)
<b>Formal + Public</b>									
Raw difference	0.101*** (0.020)	0.104*** (0.017)	0.096*** (0.016)	0.072*** (0.014)	0.041*** (0.012)	0.013 (0.009)	-0.011 (0.009)	-0.031*** (0.010)	-0.069*** (0.016)
Characteristics	0.167*** (0.022)	0.157*** (0.020)	0.142*** (0.019)	0.120*** (0.017)	0.084*** (0.014)	0.053*** (0.013)	0.033*** (0.012)	0.024* (0.014)	0.025 (0.024)
Coefficients	-0.066*** (0.017)	-0.053*** (0.017)	-0.046** (0.019)	-0.048*** (0.018)	-0.043*** (0.015)	-0.040*** (0.012)	-0.044*** (0.012)	-0.054*** (0.014)	-0.094*** (0.020)
<b>Formal + Private</b>									
Raw difference	-0.095*** (0.011)	-0.125*** (0.009)	-0.149*** (0.011)	-0.167*** (0.013)	-0.179*** (0.015)	-0.190*** (0.017)	-0.205*** (0.019)	-0.222*** (0.022)	-0.253*** (0.026)
Characteristics	-0.038*** (0.012)	-0.044*** (0.011)	-0.046*** (0.011)	-0.045*** (0.012)	-0.043*** (0.013)	-0.041*** (0.015)	-0.039*** (0.017)	-0.040*** (0.020)	-0.051*** (0.025)
Coefficients	-0.058*** (0.009)	-0.081*** (0.009)	-0.103*** (0.010)	-0.122*** (0.010)	-0.135*** (0.011)	-0.149*** (0.013)	-0.166*** (0.015)	-0.182*** (0.017)	-0.202*** (0.022)

The variance has been estimated by bootstrapping the results 100 times  
Standard errors in parentheses  
\*\*\* p<0.01, \*\* p<0.05, \* p<0.1  
Elaborated by the author

Table A.5. Estimation using Melly (2006) using as dependent variables the logarithm of average wages

	Quantile regression								
	10th	20th	30th	40th	50th	60th	70th	80th	90th
<b>Informal + Public</b>									
Raw difference	-0.589*** (0.058)	-0.607*** (0.038)	-0.560*** (0.051)	-0.466*** (0.044)	-0.423*** (0.036)	-0.413*** (0.033)	-0.410*** (0.039)	-0.386*** (0.040)	-0.437*** (0.047)
Characteristics	-0.131** (0.085)	-0.127*** (0.056)	-0.114*** (0.046)	-0.094*** (0.042)	-0.094*** (0.038)	-0.088*** (0.036)	-0.097*** (0.038)	-0.099** (0.041)	-0.111** (0.048)
Coefficients	-0.458*** (0.062)	-0.480*** (0.036)	-0.446*** (0.029)	-0.372*** (0.030)	-0.329*** (0.027)	-0.325*** (0.034)	-0.313*** (0.039)	-0.287*** (0.044)	-0.326*** (0.059)
<b>Informal + Private</b>									
Raw difference	-0.441*** (0.033)	-0.385*** (0.030)	-0.285*** (0.022)	-0.222*** (0.021)	-0.161*** (0.023)	-0.106*** (0.024)	-0.097*** (0.021)	-0.121*** (0.018)	-0.187*** (0.018)
Characteristics	-0.080*** (0.026)	-0.055*** (0.022)	0.001 (0.018)	0.038*** (0.019)	0.072*** (0.021)	0.111*** (0.022)	0.122*** (0.022)	0.102*** (0.021)	0.066*** (0.025)
Coefficients	-0.361*** (0.022)	-0.331*** (0.014)	-0.286*** (0.010)	-0.260*** (0.010)	-0.232*** (0.011)	-0.217*** (0.013)	-0.219*** (0.015)	-0.223*** (0.015)	-0.254*** (0.014)
<b>Formal + Public</b>									
Raw difference	0.051*** (0.014)	0.053*** (0.012)	0.030*** (0.009)	0.005 (0.008)	-0.014** (0.007)	-0.035*** (0.007)	-0.062*** (0.009)	-0.091*** (0.013)	-0.123*** (0.017)
Characteristics	0.128*** (0.016)	0.124*** (0.014)	0.088*** (0.011)	0.060*** (0.009)	0.042*** (0.011)	0.030*** (0.012)	0.022 (0.013)	0.011 (0.017)	0.018 (0.024)
Coefficients	-0.077*** (0.016)	-0.071*** (0.017)	-0.059*** (0.015)	-0.055*** (0.012)	-0.056*** (0.011)	-0.066*** (0.012)	-0.084*** (0.015)	-0.103*** (0.018)	-0.141*** (0.023)
<b>Formal + Private</b>									
Raw difference	-0.173*** (0.036)	-0.158*** (0.024)	-0.169*** (0.017)	-0.188*** (0.015)	-0.203*** (0.014)	-0.219*** (0.015)	-0.236*** (0.017)	-0.254*** (0.020)	-0.275*** (0.024)
Characteristics	0.022 (0.028)	-0.009 (0.019)	-0.023* (0.015)	-0.033*** (0.014)	-0.041*** (0.014)	-0.045*** (0.015)	-0.049*** (0.016)	-0.051*** (0.018)	-0.047*** (0.022)
Coefficients	-0.196*** (0.024)	-0.150*** (0.017)	-0.146*** (0.013)	-0.155*** (0.011)	-0.162*** (0.011)	-0.174*** (0.011)	-0.187*** (0.012)	-0.203*** (0.014)	-0.228*** (0.017)

The variance has been estimated by bootstrapping the results 100 times  
Standard errors in parentheses  
\*\*\* p<0.01, \*\* p<0.05, \* p<0.1  
Elaborated by the author

Table A.6. Estimation using Dinardo, Fortin and Lemieux (2009)

	Quantile regression									
	Mean	10th	20th	30th	40th	50th	60th	70th	80th	90th
<b>Informal + Public</b>										
Women	1.655*** (0.0287)	0.872*** (0.0512)	1.189*** (0.0393)	1.372*** (0.0337)	1.536*** (0.0302)	1.619*** (0.0303)	1.785*** (0.0324)	1.961*** (0.0351)	2.152*** (0.0372)	2.397*** (0.0436)
Men	1.941*** (0.0303)	1.264*** (0.0363)	1.468*** (0.0315)	1.587*** (0.0317)	1.749*** (0.0322)	1.848*** (0.0338)	1.988*** (0.0374)	2.195*** (0.0453)	2.443*** (0.0488)	2.687*** (0.0555)
Total difference	-0.286*** (0.0417)	-0.392*** (0.0627)	-0.279*** (0.0503)	-0.215*** (0.0463)	-0.214*** (0.0441)	-0.229*** (0.0453)	-0.203*** (0.0495)	-0.234*** (0.0573)	-0.291*** (0.0613)	-0.291*** (0.0706)
Explained	-0.666*** (0.169)	-0.471*** (0.0800)	-0.384*** (0.102)	-0.383*** (0.142)	-0.472** (0.220)	-0.672** (0.307)	-0.833* (0.460)	-1.107*** (0.284)	-0.990*** (0.235)	-1.148*** (0.233)
Unexplained	0.380** (0.175)	0.0783 (0.0794)	0.105 (0.105)	0.168 (0.146)	0.258 (0.224)	0.443 (0.310)	0.630 (0.463)	0.873*** (0.288)	0.699*** (0.240)	0.857*** (0.240)
<b>Explained</b>										
Total	-0.666*** (0.169)	-0.471*** (0.0800)	-0.384*** (0.102)	-0.383*** (0.142)	-0.472** (0.220)	-0.672** (0.307)	-0.833* (0.460)	-1.107*** (0.284)	-0.990*** (0.235)	-1.148*** (0.233)
Pure Explained	-0.368** (0.169)	-0.140 (0.0871)	-0.208** (0.0974)	-0.189** (0.0812)	-0.229** (0.0981)	-0.116 (0.0812)	-0.178* (0.102)	-0.280** (0.133)	-0.408** (0.192)	-0.709** (0.327)
Specif. Err.	-0.298*** (0.0788)	-0.331*** (0.0667)	-0.176*** (0.0603)	-0.194** (0.0829)	-0.243* (0.135)	-0.556** (0.278)	-0.655* (0.398)	-0.827*** (0.198)	-0.582** (0.229)	-0.439 (0.297)
<b>Unexplained</b>										
Total	0.380** (0.175)	0.0783 (0.0794)	0.105 (0.105)	0.168 (0.146)	0.258 (0.224)	0.443 (0.310)	0.630 (0.463)	0.873*** (0.288)	0.699*** (0.240)	0.857*** (0.240)
Reweight. Err.	0.282 (0.173)	0.114* (0.0684)	0.189* (0.101)	0.272* (0.144)	0.413* (0.222)	0.529* (0.307)	0.713 (0.459)	0.425 (0.284)	0.188 (0.226)	0.192 (0.224)
Pure Unexplained	0.0977** (0.0428)	-0.0358 (0.0565)	-0.0840 (0.0511)	-0.104* (0.0575)	-0.155** (0.0677)	-0.0859 (0.0759)	-0.0830 (0.0932)	0.448*** (0.0619)	0.511*** (0.0719)	0.665*** (0.0770)

Standard errors in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

	Quantile regression									
	Mean	10th	20th	30th	40th	50th	60th	70th	80th	90th
<b>Informal + Private</b>										
Women	1.369*** (0.00982)	0.599*** (0.0203)	0.921*** (0.0151)	1.138*** (0.0114)	1.289*** (0.00987)	1.417*** (0.00921)	1.535*** (0.00884)	1.652*** (0.00920)	1.816*** (0.0108)	2.062*** (0.0161)
Men	1.575*** (0.00710)	0.819*** (0.0156)	1.148*** (0.0106)	1.335*** (0.00785)	1.478*** (0.00705)	1.606*** (0.00664)	1.725*** (0.00680)	1.843*** (0.00714)	2.014*** (0.00868)	2.299*** (0.0127)
Total difference	-0.206*** (0.0121)	-0.220*** (0.0256)	-0.227*** (0.0185)	-0.196*** (0.0138)	-0.189*** (0.0121)	-0.189*** (0.0114)	-0.190*** (0.0112)	-0.191*** (0.0116)	-0.198*** (0.0139)	-0.237*** (0.0205)
Explained	-0.0994** (0.0392)	-0.0107 (0.0613)	-0.0541 (0.0423)	-0.0679** (0.0314)	-0.0633** (0.0281)	-0.0657** (0.0265)	-0.0638** (0.0282)	-0.0895** (0.0384)	-0.121** (0.0536)	-0.226*** (0.0783)
Unexplained	-0.107** (0.0435)	-0.209*** (0.0717)	-0.173*** (0.0492)	-0.128*** (0.0362)	-0.126*** (0.0320)	-0.123*** (0.0299)	-0.126*** (0.0317)	-0.101** (0.0424)	-0.0773 (0.0584)	-0.0112 (0.0850)
<b>Explained</b>										
Total	-0.0994** (0.0392)	-0.0107 (0.0613)	-0.0541 (0.0423)	-0.0679** (0.0314)	-0.0633** (0.0281)	-0.0657** (0.0265)	-0.0638** (0.0282)	-0.0895** (0.0384)	-0.121** (0.0536)	-0.226*** (0.0783)
Pure Explained	-0.0581* (0.0298)	0.0335 (0.0304)	-0.00432 (0.0259)	-0.0247 (0.0223)	-0.0301 (0.0211)	-0.0484** (0.0222)	-0.0491** (0.0228)	-0.0664*** (0.0252)	-0.0820*** (0.0307)	-0.128*** (0.0467)
Specif. Err.	-0.0413** (0.0201)	-0.0442 (0.0409)	-0.0498* (0.0272)	-0.0431** (0.0189)	-0.0333* (0.0177)	-0.0173 (0.0157)	-0.0147 (0.0186)	-0.0232 (0.0252)	-0.0386 (0.0353)	-0.0976* (0.0546)
<b>Unexplained</b>										
Total	-0.107** (0.0435)	-0.209*** (0.0717)	-0.173*** (0.0492)	-0.128*** (0.0362)	-0.126*** (0.0320)	-0.123*** (0.0299)	-0.126*** (0.0317)	-0.101** (0.0424)	-0.0773 (0.0584)	-0.0112 (0.0850)
Reweight. Err.	0.0844** (0.0330)	0.0966** (0.0470)	0.0983** (0.0427)	0.0635** (0.0308)	0.0531** (0.0268)	0.0636** (0.0263)	0.0638*** (0.0236)	0.0641** (0.0299)	0.0860** (0.0411)	0.0964* (0.0566)
Pure Unexplained	-0.191*** (0.0443)	-0.306*** (0.102)	-0.271*** (0.0720)	-0.192*** (0.0476)	-0.179*** (0.0387)	-0.187*** (0.0345)	-0.190*** (0.0295)	-0.165*** (0.0392)	-0.163*** (0.0494)	-0.108 (0.0670)

Standard errors in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1



	Quantile regression									
	Mean	10th	20th	30th	40th	50th	60th	70th	80th	90th
<b>Formal + Public</b>										
Women	2.597*** (0.0104)	1.890*** (0.0221)	2.197*** (0.0182)	2.370*** (0.0168)	2.534*** (0.0144)	2.674*** (0.0113)	2.766*** (0.00998)	2.858*** (0.00989)	2.968*** (0.0109)	3.144*** (0.0148)
Men	2.569*** (0.0123)	1.827*** (0.0200)	2.068*** (0.0179)	2.257*** (0.0186)	2.451*** (0.0186)	2.624*** (0.0159)	2.750*** (0.0129)	2.861*** (0.0125)	2.997*** (0.0140)	3.207*** (0.0195)
Total difference	0.0282* (0.0161)	0.0631** (0.0298)	0.129*** (0.0255)	0.113*** (0.0251)	0.0826*** (0.0235)	0.0503*** (0.0195)	0.0166 (0.0163)	-0.00328 (0.0160)	-0.0290 (0.0177)	-0.0636*** (0.0245)
Explained	0.0275 (0.0382)	0.0716*** (0.0230)	0.182 (0.118)	0.0810 (0.0721)	0.0470 (0.0499)	0.0215 (0.0326)	0.00292 (0.0231)	-0.0127 (0.0193)	-0.0393** (0.0187)	-0.0740*** (0.0246)
Unexplained	0.000664 (0.0424)	-0.00850 (0.0312)	-0.0531 (0.122)	0.0320 (0.0770)	0.0356 (0.0558)	0.0288 (0.0387)	0.0137 (0.0289)	0.00937 (0.0257)	0.0103 (0.0264)	0.0104 (0.0359)
<b>Explained</b>										
Total	0.0275 (0.0382)	0.0716*** (0.0230)	0.182 (0.118)	0.0810 (0.0721)	0.0470 (0.0499)	0.0215 (0.0326)	0.00292 (0.0231)	-0.0127 (0.0193)	-0.0393** (0.0187)	-0.0740*** (0.0246)
Pure Explained	0.103*** (0.0342)	0.287** (0.135)	0.198*** (0.0729)	0.154*** (0.0515)	0.124*** (0.0354)	0.0755*** (0.0232)	0.0531*** (0.0173)	0.0404*** (0.0147)	0.0299** (0.0138)	0.0280* (0.0157)
Specif. Err.	-0.0760*** (0.00666)	-0.215 (0.145)	-0.0158 (0.0477)	-0.0734*** (0.0229)	-0.0774*** (0.0169)	-0.0540*** (0.0108)	-0.0502*** (0.00702)	-0.0530*** (0.00663)	-0.0693*** (0.00816)	-0.102*** (0.0150)
<b>Unexplained</b>										
Total	0.000664 (0.0424)	-0.00850 (0.0312)	-0.0531 (0.122)	0.0320 (0.0770)	0.0356 (0.0558)	0.0288 (0.0387)	0.0137 (0.0289)	0.00937 (0.0257)	0.0103 (0.0264)	0.0104 (0.0359)
Reweight. Err.	0.000366 (0.0398)	-0.00172 (0.0191)	0.0233 (0.120)	0.00727 (0.0741)	0.00313 (0.0516)	-0.000708 (0.0339)	-0.00331 (0.0242)	-0.00389 (0.0199)	-0.00427 (0.0183)	-0.00701 (0.0213)
Pure Unexplained	0.000298 (0.0159)	-0.00678 (0.0252)	-0.0764** (0.0316)	0.0248 (0.0244)	0.0324 (0.0218)	0.0295 (0.0185)	0.0170 (0.0160)	0.0133 (0.0166)	0.0146 (0.0197)	0.0174 (0.0303)

Standard errors in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

	Quantile regression									
	Mean	10th	20th	30th	40th	50th	60th	70th	80th	90th
<b>Formal + Private</b>										
Women	2.105*** (0.0128)	1.502*** (0.0110)	1.629*** (0.0108)	1.744*** (0.0112)	1.849*** (0.0123)	1.972*** (0.0143)	2.123*** (0.0168)	2.288*** (0.0199)	2.546*** (0.0259)	2.964*** (0.0315)
Men	2.284*** (0.0101)	1.582*** (0.00922)	1.738*** (0.00951)	1.883*** (0.0104)	2.032*** (0.0112)	2.178*** (0.0117)	2.328*** (0.0127)	2.523*** (0.0149)	2.759*** (0.0172)	3.162*** (0.0252)
Total difference	-0.179*** (0.0163)	-0.0806*** (0.0143)	-0.109*** (0.0144)	-0.139*** (0.0153)	-0.184*** (0.0166)	-0.207*** (0.0185)	-0.205*** (0.0211)	-0.235*** (0.0248)	-0.212*** (0.0311)	-0.198*** (0.0403)
Explained	-0.0689* (0.0372)	0.0159 (0.0332)	0.00755 (0.0289)	-0.00211 (0.0294)	-0.00908 (0.0337)	-0.0329 (0.0451)	-0.0800 (0.0564)	-0.125* (0.0665)	-0.283*** (0.0949)	-0.135*** (0.0412)
Unexplained	-0.110** (0.0441)	-0.0965** (0.0388)	-0.117*** (0.0349)	-0.136*** (0.0360)	-0.175*** (0.0410)	-0.174*** (0.0534)	-0.125* (0.0654)	-0.110 (0.0757)	0.0706 (0.105)	-0.0624 (0.0571)
<b>Explained</b>										
Total	-0.0689* (0.0372)	0.0159 (0.0332)	0.00755 (0.0289)	-0.00211 (0.0294)	-0.00908 (0.0337)	-0.0329 (0.0451)	-0.0800 (0.0564)	-0.125* (0.0665)	-0.283*** (0.0949)	-0.135*** (0.0412)
Pure Explained	0.0611*** (0.0226)	0.0461*** (0.0178)	0.0611*** (0.0165)	0.0617*** (0.0178)	0.0805*** (0.0192)	0.0870*** (0.0223)	0.0793*** (0.0262)	0.0689** (0.0274)	0.0605* (0.0360)	-0.00534 (0.0416)
Specif. Err.	-0.130*** (0.0277)	-0.0302 (0.0227)	-0.0536** (0.0215)	-0.0638*** (0.0219)	-0.0896*** (0.0280)	-0.120*** (0.0387)	-0.159*** (0.0482)	-0.194*** (0.0532)	-0.344*** (0.0795)	-0.130*** (0.0457)
<b>Unexplained</b>										
Total	-0.110** (0.0441)	-0.0965** (0.0388)	-0.117*** (0.0349)	-0.136*** (0.0360)	-0.175*** (0.0410)	-0.174*** (0.0534)	-0.125* (0.0654)	-0.110 (0.0757)	0.0706 (0.105)	-0.0624 (0.0571)
Reweight. Err.	-0.0295 (0.0329)	-0.00189 (0.0243)	-0.000205 (0.0229)	-0.00515 (0.0219)	-0.0159 (0.0237)	-0.0170 (0.0324)	-0.0345 (0.0415)	-0.0545 (0.0626)	-0.0861 (0.0876)	-0.0391 (0.0458)
Pure Unexplained	-0.0806* (0.0459)	-0.0946*** (0.0365)	-0.117*** (0.0357)	-0.131*** (0.0353)	-0.159*** (0.0409)	-0.157*** (0.0547)	-0.0901 (0.0661)	-0.0551 (0.0801)	0.157 (0.110)	-0.0232 (0.0751)

Standard errors in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Elaborated by the author